DOCUMENT RESUME

ED 152 781 TM 006 536

AUTHOR Anderson, Scarvia B.
TITLE Issues Related to Test Use.
PUB DATE 5 Apr 77
NOTE 8p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New York, New York, April 5-7, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Educational Testing; Elementary Secondary Education; *Evaluators; Manuals; Post Secondary Education; Psychometrics; Qualifications; *Standards; Testing; *Testing Problems; *Test Interpretation; Tests; Test Selection; Test Validity
IDENTIFIERS *Teacher Developed Tests; *Testing Industry; Test Manuals

ABSTRACT

Several issues are related to the use of educational tests. First, test users must be able to choose appropriate tests, interpret scores, and make decisions based on scores. In the field of educational testing, few test users have adequate training in these areas. Second, test makers must clearly specify directions for administration, allowable test uses, and appropriate performance interpretations. Third, the performance of test makers in relationship to test users should be closely monitored by both professions. The final issue is related to the use of locally-developed tests. Since decisions made on the basis of these tests have a profound influence on the lives of students, measurement experts should take deliberate actions to promote adequate practices in this area through training and consultation. (BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED152781

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ISSUES RELATED TO TEST USE*

Scarvia B. Anderson
Educational Testing Service

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Scarvia B. Anderson
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

To paraphrase the eminent sociologist, Melvin Tumin (1969): It is axiomatic that when a number of parties are involved in any social enterprise, and when problems arise, each party will lay maximum blame on the others and will assume only minimum blame, if any, for itself.

We have certainly seen this axiom proved in the current controversy about educational testing. Test takers blame test makers, attributing any deficiencies in their own performance to the characteristics of the tests. Test makers blame test users, attributing any deficiencies in measurement to the application and interpretation. Test users blame both test takers and test makers--test takers for involving them in the process in the first place, and test makers for not involving themselves in the process sufficiently.

Professor Tumin goes on to say: As a corollary, it follows that the official verdict of guilt for the situation will be imposed on that party who is weakest or least able to fend off the imposition of the official stigma. In the case of testing, this second proposition seems to break down. The majority of attacks are levied on the party that seems the strongest--the national testing organizations. At least they are the most organized and visible.

Of course, the alternative for test users of blaming poor test performance on the educational system, of which they are a part, or test

---

* Attacks on Testing: The Issues, symposium, National Council on Measurement in Education, New York City, April 5, 1977

TM006 536

misuse on their own measurement incompetencies is unthinkable. Just as is the alternative for test takers of blaming poor test performance on their own inattention or inadequacies. Recently and occasionally a few test takers, and their parents, have held local schools at least partly responsible for poor showings on tests. But, in the main, it has been the national testing organizations who have had to take the brunt of criticisms about testing.

This is not to say that these test makers are guiltless. However, test takers have their own houses to see to. And test users--who are supposed to be the focus of my remarks--cannot shove all of the blame for their improprieties on others. Unless it is in the area of what is expected of them. Let's review their case there:

The APA-AERA-NCME Standards for Educational & Psychological Tests (1974) define a "test user" as "one who chooses tests, interprets scores, or makes decisions based on scores." According to these Standards, the following, among others, are considered ESSENTIAL competencies for test users; test users should

* have an understanding of psychological or educational measurement and of validation and other research,
* consider more than one variable for assessment, and the assessment of any given variable by more than one method,
* relate the history of research and development of an existing test to the intended use of the instrument, and
* recognize that estimates of reliability do not indicate criterion-related validity.

A rather heavy load to place on the typical users of educational tests: teachers, school administrators, counselors, admissions officers, officials of agencies that control educational funds. The fact is that few of them have time--or have had the time or opportunity in the past--for more than cursory study of psychometric theory and procedures. Most

of them are exceptionally busy pursuing success and proficiency in remedial reading, zero-based budgeting, occupational guidance, scholarship services, or development of RFPs. Furthermore, those who also serve in an instructional capacity have to devote many hours to developing, scoring, and interpreting their own tests, without the added nuisance of dealing with tests developed by others. (It is well to remind ourselves occasionally that by far the largest and most pervasive use of educational tests in the United States is by local people who make tests—the same local people who also make the rules for using the tests to assign grades and for other educational decisions.)

The Standards hint at the possibility that test specialists' expectations with respect to the sophistication of many test users may be unrealistic. For example, on page 59, test users with limited technical training are advised to "seek refresher training or work under the guidance of another test user whose training is adequate." And on page 68, this admittedly unanswered question is posed: "Whether ...a person who lacks the training necessary for the interpretation of scores should be given that training or should be given only the interpretations of scores."

A major issue in test use, then, is how much psychometric sophistication most users of most educational tests need to have. It seems time to revisit the Standards on this point and focus specifically on tests designed for educational information and decisions. You will recall that the 1974 Standards were developed to encompass all kinds of tests, including tests administered in clinical and industrial settings. While many of the same principles apply, the nature of the tests used in such different settings, as well as the characteristics of the users, can be expected to vary considerably.

I submit that for most educational purposes, it is not necessary to develop measures that require an Eric Gardner, Bob Ebel, or Ellis Page to understand and use intelligently. And I am not advancing this proposition just for the sake of expediency, recognizing the frequently limited competencies of present users of educational tests. I am also advancing it in the name of the central concept of our profession: validity. A good case can be made for a positive relationship between the validity of educational measurement and the straightforwardness of its conception ("straightforward" in the sense of bearing a direct and observable relationship to educational goals and requirements); and a corresponding negative relationship between validity and the number or size of the inferential leaps required to make measurement-derived decisions. It follows from this line of reasoning that the technical competencies of most users would not need to extend beyond the ability to read a test publication that spells out clearly the uses for which a test is intended and provides interpretations of performance oriented directly toward those uses. If users are the test administrators or scorers, then directions for those activities would have to be prescribed for them too.

It also follows that the requirement of clear specification of directions, allowable test uses, and appropriate performance interpretations places a heavy responsibility on test makers. However, it is a responsibility they have already accepted in principle, although the execution (especially with respect to clarity) has not always been notable. It would be up to the test maker earlier to determine whether the educational measures were in fact amenable to this kind of treatment that assumes relatively little

sophistication in tests and measurements on the part of potential users. I have suggested--and believe--that most measures of educational ability, achievement, interests, and so on can fit into this category. But we must allow for a few good measures that might not; and in those cases, precautions would have to be taken that they did not get into the wrong hands or that people unprepared to use them were alerted to seek assistance. Establishing workable procedures for such limitations or cautions is a second issue related to test use.

A third issue is even touchier. What if test makers do not delineate acceptable measurement purposes and performance interpretations? Or, if test makers have done all they can to inform test users and lead them to rational uses, what if test users deliberately or neglectfully disregard that advice? Up to now, any rebukes for real or perceived shortcomings in national tests have been rather haphazard. They have taken three major forms: professional critiques in the Mental Measurements Yearbooks and other publications--often appearing years after a measure was made available and seldom read by potential users of the measure; journalistic attacks, sometimes sound but sometimes self-serving; and complaints from test takers and users. The last have probably been the most persuasive force for change and improvement in national tests and services, especially if the users accompanied their criticisms with threats to stop using the tests or services if changes and improvements weren't made.

Sanctions against test users who were in fact misusers have been even more haphazard but sometimes more drastic. The courts have applied the more drastic ones. Otherwise, some test makers have attempted to follow up on flagrant misuses that were brought forcefully to their attention. However, it has not been practical or politically feasible for even the most professionally dedicated of the test makers to ferret out the majority

of misuses. And even when they do follow up, the legal basis for how far they can go in withdrawing instruments and services from recidivistic misusers is not thoroughly established.

From the standpoint of my personal political persuasion, it is tempting to suggest that test users and test takers, with their obvious economic clout, be left alone to deal with inadequate performance or misconduct by test makers. However, here, and even more so in the case of test misuse, a laissez-faire approach does not seem adequate. The often legitimate concerns of the EEOC and other federal agencies for test practices in business and industry have so often had a fallout of bureaucratic meddling that it behooves us in educational testing to do whatever is necessary to make further government surveillance unnecessary. Therefore, the sensible approach appears to be systematic and timely monitoring of testing services and uses <u>within</u> the educational and testing professions. Otherwise, we will have abdicated our responsibilities to the federal government or, worse, to the vagaries of a Ralph Nader.

I have talked about three issues related to use of national tests that we as a profession need to give increased attention to: competencies we can reasonably expect test users to have, any limitations required on distribution and use of educational tests related to test users' competencies, and effective monitoring of the performance of test makers in relationship to test users. Let me return to a fourth and--for these remarks--final issue. The majority of test use occurs with locally--not nationally--developed tests. Some will say, "Yes, but those uses aren't so critical; they don't figure in decisions to deny college admissions or relegate students to classes for the educable mentally retarded." However, even those examples are not universally accurate; some such decisions <u>are</u> made

using local instruments. More important: What about the cumulative effects on students' self-concepts of feedback from locally developed examinations and quizzes? What about the grades students earn which loom large in decisions about promotion, placement, graduation, and indeed college admissions and employment—and are based to a large extent on performance on local tests? What about innovative programs that are judged deserving of rapid and enthusiastic dissemination to other districts or institutions on the basis of evaluations employing locally developed tests?

Influencing tens of thousands of local test maker-users is a formidable task compared with keeping up with twenty or so national organizations responsible for tests designed for wide-scale educational use. However, when we consider the profound influence of local tests on the lives of students and instructors and on the workings of institutions, we must as a profession of measurement experts in the service of education take some deliberate and responsible actions to help promote adequate practices in this arena—through training, consultation, and perhaps illumination. It's not a question of we and they; it's us.

References

American Psychological Association. Standards for educational and psychological tests. Washington, D.C.: Author, 1974. (Prepared by a joint committee of the APA, AERA, and NCME.)

Tumin, M. The family and community: What are their roles in the educational process? In Untangling the tangled web of education. Princeton, N.J.: Educational Testing Service, 1969.